Abe Jacobson

ajacobson@lanl.gov

15 February 2000

Using the augmented rf database tool

(1) Introduction

This note describes a database tool for FORTE workers to access and use the second-generation rf database for TATR (50 MegaSample/sec) data. This is an improvement over the rf database presented in the "lgu" files. The second generation is slower to use, has fewer events satisfying the various qualifying tests, but in the end gives significantly more information, and more correct information, for the fewer events which do qualify.

(2) The form of the database

The new database is in an unformatted storage and is based on the analysis code totaldechirp.pro. Therefore the database files are called "tdu" files. As of this writing, here are the files (and sizes, in kilobytes):

ls -1 -s /n/projects/sat/idl/tdu

total 1013084

5520 199709.tdu

6764 199710.tdu

20324 199711.tdu

19604 199712.tdu

19932 199801.tdu

56108 199802.tdu

92488 199803.tdu

70616 199804.tdu

89540 199805.tdu

152108 199806.tdu

99228 199807.tdu

94288 199808.tdu

57596 199809.tdu

48856 199810.tdu

97976 199811.tdu

82136 199908.tdu

There is one file for each month analyzed so far. The files are very big (each is tens of megabytes)! When you actually access this data, be prepared to wait a few minutes for the data to come to you. And don't even dream of using a small-RAM box.

You may access a single month's file (1998 June) using the IDL function readtdu.pro. For example, you might enter the command

data=readtdu('199806')

and it will immediately bark back at you

# of one-freq OK events: 142823

# of two-freq OK events: 27150

This gives the number of qualifying events in this file with single-TATR frequency (both TATRs at the same frequency) and with two-TATR frequencies (the TATRs at different frequencies).

What is *not* so immediate is the return of the data itself.

The data possibilities for these two cases are a bit different, so there are different structures for each of these two cases, all bundled into the capsule structure "data". Data includes at the top level:

data.i1 = # of single-freq events

data.i2 = # of dual-freq events

data.g1 = array of formally identical structures (count=data.i1) containing housekeeping info for each single-freq event

data.g2 = array of formally identical structures (count=data.i2) containing housekeeping info for each dual-freq event.

data.a = array of formally identical structures (count=data.i1) containing computed information from TATRA data for each single-freq event.

data.b = array of formally identical structures (count=data.i1) containing computed information from TATRB data for each single-freq event.

data.l = array of formally identical structures (count=data.i2) containing computed information from TATR data at the low-band frequency for each dual-freq event.

data.h = array of formally identical structures (count=data.i2) containing computed information from TATR data at the high-band frequency for each dual-freq event.

(3) Database contents for a single-frequency event

Now let's illustrate the contents of these structures for a single-freq event. First, the "housekeeping info a single-freq event is as follows:

IDL> help,data.g1(0),/structure

** Structure <82a9924>, 29 tags, length=84, refs=2:

DASFILE STRING '19980605_173618'

NTH LONG 22

LON FLOAT -139.446

LAT FLOAT 10.4676

ALT FLOAT 802.680

RUNID LONG 581056757

EVTNUM LONG 21

LSEC LONG 581056932

DSEC DOUBLE 0.13574070

FREQA FLOAT 38.0000

FREQB FLOAT 38.0000

PREAMP1 BYTE 1

PREAMP2 BYTE 1

BUANTPWR BYTE 0

ANTTYPE BYTE Array[2]

TRIGSRCCODE BYTE 0

NBEFORE LONG 5099

TATRAPOWER BYTE 1

TATRAANTENNA BYTE 1

TATRBPOWER BYTE 1

TATRBANTENNA BYTE 0

TRIGAPOWER BYTE 1

TRIGATDL BYTE 0

TRIGACOINCHS BYTE 5

TRIGACOINWIN LONG 162

TRIGBPOWER BYTE 1

TRIGBTDL BYTE 0

TRIGBCOINCHS BYTE 5

TRIGBCOINWIN LONG 20

IDL>

All of these items are familiar from the "lgu" files containing the original rf database.

Next, the TATRA data for this single-freq data is as follows:

IDL> help,data.a(0),/structure

** Structure <82aa2e4>, 27 tags, length=340, refs=2:

TEC0A FLOAT 4.72229

TECOPTA FLOAT 4.72229

QUART100A FLOAT 0.00000

TEC0OPTA FLOAT 0.00000

SIGRATA FLOAT 0.00000

FCEBOPTA FLOAT -99.0000

TIPPSPLITA FLOAT 27.6800

TIPPCORA FLOAT 0.390261

SNR2A FLOAT 204.725

WHALFA FLOAT 3.36000

TIPPSPLITAA FLOAT Array[8]

TIPPCORAA FLOAT Array[8]

SNR2AA FLOAT Array[8]

WHALFAA FLOAT Array[8]

FAA FLOAT Array[8]

CONTRASTA FLOAT 1199.87

PPEAKA FLOAT 5.38370e-07

PAVGA FLOAT 1.05372e-08

COHERENCEA BYTE 0

NSUBA BYTE 12

TSMMAXA FLOAT 140.000

PSMMAXA FLOAT 4.65287e-07

TSMA BYTE Array[50]

PSMA BYTE Array[50]

FREQRADARA FLOAT 0.00000

POWRADARA FLOAT 0.00000

RADARA STRING ' radar:0'

IDL>

tec0a = the simple determination of TEC, without allowing a quartic contribution. It is similar to the TEC determined in "lgu" database.

tecopta = improved estimate of TEC, allowing the quartic to be fit. If the quartic contribution is not fitted with significance, then tecopta=tec0a remains the value of TEC.

quart100a = quartic delay (μs) scaled to 100 MHz. If this is pinned at zero, then it wasn't fitted with adequate significance.

sigrata = significance ratio for quart100a determination. Prudence dictates: If sigrata is <3, don't bother to use the quartic contribution, and stick with quadratic-only determination (tec0a) of TEC.

fcebopta = fitted value of $|f_{ce}X\cos(\beta)|$, where f_{ce} is the electron gyrofrequency, and β is the angle between geomagnetic field and line-of-sight. If this is -99, a fit was either not justified, or was attempted but was not significant.

tippsplita = TIPP splitting (μs) for the uppermost 8 MHz of band. This avoids the refractive-bending-generated suppression of splitting at lowest frequencies. If there is no significance to the fit, tippsplita=0.

tippcora = power autocorrelation function secondary-peak amplitude. Is a measure of how strong the TIPP second pulse is relative to first pulse. Don't sweat this parameter.

snr2a = snr of power autocorrelation function secondary peak (see lgu database). Hint: To be prudent, don't use the TIPP splitting unless snr2a >8.

whalfa = 1/e halfwidth (μs) for the primary peak of the power autocorrelation function.

tippsplitaa,tippcoraa,snr2aa, and whalfaa: arrays of the above for separate determinations of TIPP in each of eight different eighths of the passband. This is good for tracing the effects of raybending on the TIPP splitting. This is attempted only for high TEC; otherwise left zeroes. Not for usual work.

faa=frequency array for the above arrays. This is weighted by the power-weighted centroid within each eighth of the passband. This is attempted only for high TEC; otherwise left zeroes.

contrasta = peak power divided by median power (see lgu database).

ppeaka = peak square of electric field ($(V/m)^2$).

pavga = averaged square of electric field ($(V/m)^2$).

coherencea = byte indicator of whether the pulse is coherent or not (=1 means coherent; =0 means incoherent.) Uses a fading test applied to brightest 40 μs.

nsub = number of non-zeroed subsamples for time record of square of electric field ($(V/m)^2$).

tsmmaxa = maximum value (μs) of subsampled time series times.

psmmaxa = maximum value ($(V/m)^2$) of subsampled time series of squared electric field.

tsm = byte-scaled subsample time array (not sorted)

psm = corresponding byte-scaled subsample d array of squared electric field.

For example, to print the times in this subsample array, do the following:

```
IDL> print,float(data.a(0).tsma)*(data.a(0).tsmmaxa/255.)

31.8431 140.000 0.00000 28.0000 136.157

108.157 19.7647 7.68628 11.5294 127.922

116.392 124.078 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000
```

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

IDL>

Similarly, to get the series of squared electric field, do the following:

IDL>
print,float(data.a(0).psma)*(data.a(0).psmmaxa/255.)

0.00000 9.12327e-09 4.92657e-08 6.56876e-08 1.11304e-07

1.13129e-07 1.51446e-07 1.84290e-07 2.26257e-07 3.01068e-07

3.79528e-07 4.65287e-07 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

0.00000 0.00000 0.00000 0.00000 0.00000

IDL>

Note that there are only nsuba (=12) nontrivial values in these subsampled arrays. The number of nontrivial samples depends on the complexity of the rf envelope pulse. There will always be 50 total elements in both tsma and psma, of which nsuba will be nontrivial.

This procedure "powertime.pro" allows for easy use of the data from a single event, either single-frequency or two-frequency. Let g be either g1 (single-frequency) or g2 (two-frequency), and let d be the actual data structure (i.e., either a or b for single-frequency, or l for two-frequency). Then you can extract floating-point, rescaled, and interpolated data with powertime.pro:

This procedure takes the substructures g and d (for one event) and uses them to reconstruct the power-vs-time. At the end, outputs floating-point data with all points explicitly inserted. Only low-band power is outputted, from 1-freq or 2-freq data.

Caution: if nsub=0 at output, there is no subsampled data, and if nsub=50 at output, there is data but there is likely fragmentation due to there not being enough samples.

;

;Inputs:

; g=substructure data.gX(j) where j=specific event and X=1,2

; d=substructure data.Y(j) where j=specific event and Y=a,b,l

;

;Outputs:

; nout=number of output points

; tout=output float time array, units=microseconds

; pout=output float power array, units= (v/m)**2

; nsub=number of subsampled points (caution if nsub=50)

;

;-

pro powertime,g,d,nout,tout,pout,nsub

freqradara = frequency (MHz) of a detected radar. (=0 if there is no radar detected).

powradara = squared electric field ($(V/m)^2$) for radar signal, if any.

radara = string warning of detection of a radar signal.

= ' radar:0' if nothing, =' radar:1' if radar detected.

The TATRB data for this single-frequency file is almost identical in form to that of TATRA:

IDL> help,data.b(0),/structure

** Structure <82aaeec>, 28 tags, length=344, refs=2:

TEC0B FLOAT 4.64383

TECOPTB FLOAT 4.64383

QUART100B FLOAT 0.00000

TEC0OPTB FLOAT 0.00000

SIGRATB FLOAT 0.00000

FCEBOPTB FLOAT -99.0000

TIPPSPLITB FLOAT 27.6800

TIPPCORB FLOAT 0.322827

SNR2B FLOAT 104.316

WHALFB FLOAT 3.04000

TIPPSPLITAB FLOAT Array[8]

TIPPCORAB FLOAT Array[8]

SNR2AB FLOAT Array[8]

WHALFAB FLOAT Array[8]

FAB FLOAT Array[8]

CONTRASTB FLOAT 694.743

PPEAKB FLOAT 2.38534e-07

PAVGB FLOAT 3.89649e-09

COHERENCEB BYTE 0

NSUBB BYTE 12

TSMMAXB FLOAT 140.000

PSMMAXB FLOAT 1.78174e-07

TSMB BYTE Array[50]

PSMB BYTE Array[50]

FREQRADARB FLOAT 0.00000

POWRADARB FLOAT 0.00000

RADARB STRING ' radar:0'

POWRATAB FLOAT 2.61435

IDL>

The only additional parameter is

powratab= (TATRA integrated pulse energy)/(TATRB pulse energy); this is a robust estimate of relative power using the integral over the bright central pulse. If the waveform is diffuse (i.e., no single dominant pulse) this ratio is left =0. The ratio powratab is for careful antenna-lobe work using the data from two antennas but at the same frequency. The ratio is gotten using the same prewhitening (carrier suppression) weighs in both TATRA and TATRB, in order not to bias the power ratio. Of course, this parameter will not be computed, i.e. will be left zero, if there needed to be an unscrambling (in fql) of either TATR's data. That is because the unscrambling may not entirely correct the amplitudes.

(4) Database contents for a dual-frequency event

In the case of dual-freq data, we use the low-band data as primary, and use the high-band data primarily to strengthen the joint quadratic/quartic fit to dispersion.

The housekeeping information is identical in form to that for a single-freq event:

IDL> help,data.g2(0),/structure

**** Structure <82ab8a4>, 29 tags, length=84, refs=2:**

DASFILE STRING '19980601_003918'

NTH LONG 1

LON FLOAT -74.7657

LAT FLOAT 57.8173

ALT FLOAT 812.880

RUNID LONG 580691536

EVTNUM LONG 0

LSEC LONG 580691552

DSEC DOUBLE 0.14820810

FREQA FLOAT 38.0000

FREQB FLOAT 130.000

PREAMP1 BYTE 1

PREAMP2 BYTE 1

BUANTPWR BYTE 0

ANTTYPE BYTE Array[2]

TRIGSRCCODE BYTE 0

NBEFORE LONG 5099

TATRAPOWER BYTE 1

TATRAANTENNA BYTE 1

TATRBPOWER BYTE 1

TATRBANTENNA BYTE 0

TRIGAPOWER BYTE 1

TRIGATDL BYTE 0

TRIGACOINCHS BYTE 5

TRIGACOINWIN LONG 162

TRIGBPOWER BYTE 1

TRIGBTDL BYTE 0

TRIGBCOINCHS BYTE 5

TRIGBCOINWIN LONG 20

IDL>

The low-band data for this event is as follows:

IDL> help,data.l(0),/structure

** Structure <82ac274>, 28 tags, length=340, refs=2:

TEC0L FLOAT 4.80758

TECOPTL FLOAT 4.07830

QUART100L FLOAT 0.0478155

TEC0OPTL FLOAT 4.43753

SIGRATL FLOAT 0.00000

FCEBOPTL FLOAT -99.0000

TIPPSPLITL FLOAT 0.00000

TIPPCORL FLOAT 0.00000

SNR2L FLOAT 0.00000

WHALFL FLOAT 5.60000

TIPPSPLITAL FLOAT Array[8]

TIPPCORAL FLOAT Array[8]

SNR2AL FLOAT Array[8]

WHALFAL FLOAT Array[8]

FAL FLOAT Array[8]

CONTRASTL FLOAT 125.577

PPEAKL FLOAT 3.64462e-07

PAVGL FLOAT 7.96410e-09

COHERENCEL BYTE 0

BIFREQ BYTE 1

NSUBL BYTE 8

TSMMAXL FLOAT 99.0000

PSMMAXL FLOAT 2.80764e-07

TSML BYTE Array[50]

PSML BYTE Array[50]

FREQRADARL FLOAT 0.00000

POWRADARL FLOAT 0.00000

RADARL STRING ' radar:0'

IDL>

Here, the only addition relative to single-freq "a" data is:

bifreq = byte flag of whether the high-band frequency could participate in the fit of coordinated quadratic/quartic dispersion parameters. If bifrqeq=1, then the fit (to dispersion) was able to use the high-band data; if bifreq=0, the high-band data did not qualify to participate in the fit. The best (i.e., highest-leverage) fits to dispersion will be with bifreq=1. Note that in the example given here, the high band did participate.

One other small difference: The string flag "radarl" can be toggled by a radar's being detected in either low-band or high-band.

Note that for this event, there was a significant fit of coordinated quadratic/quartic dispersion parameters (due to sigratl > 3.0), and that as a result we have tecoptl , tec0l, and quart100l > 0.

The additional information from the high-band data is a subset of the above:

IDL> help,data.h(0),/structure

** Structure <82acaa4>, 14 tags, length=148, refs=2:

TIPPSPLITH FLOAT 0.00000

TIPPCORH FLOAT 0.00000

SNR2H FLOAT 0.00000

WHALFH FLOAT 34.4000

CONTRASTH FLOAT 20.0960

PPEAKH FLOAT 1.49105e-08

PAVGH FLOAT 8.71829e-10

NSUBH BYTE 5

TSMMAXH FLOAT 60.0000

PSMMAXH FLOAT 1.23266e-08

TSMH BYTE Array[50]

PSMH BYTE Array[50]

FREQRADARH FLOAT 0.00000

POWRADARH FLOAT 0.00000

IDL>

Note that there is not a separate string flag for radar; instead, the low-band flag radarl is toggled by the presence of a radar detection in either low-band or accompanying high-band.

(5) Examples of how to use the database

(5a) Parse for radars

Radars are defined in this database as short, narrowband pulses with relatively steady power within the "on" duration. Radars can do incredible mischief to statistics about pulse shape of lightning signals etc.

The algorithm is intentionally generous, so there'll be lots of false positives (just like a polygraph!)

To limit false positives in your search, ratchet up the power criterion (via powradar) in your data search, or limit to certain regions geographically. Be sure to look at the actual DAS data after you've gotten a pointer to it.

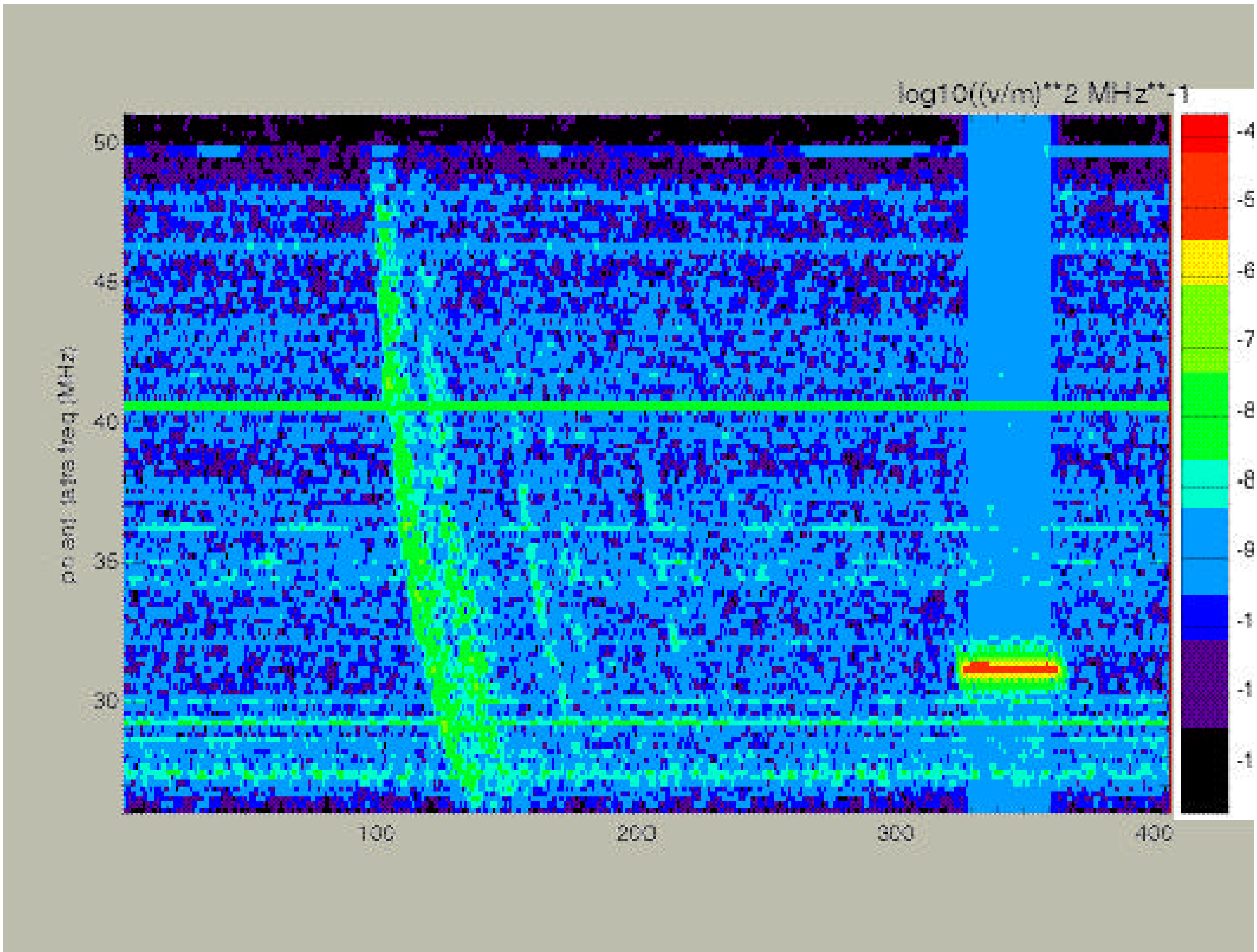
Suppose you want to find all the events that are corrupted by the presence of a radar signal with an electric field exceeding the very unusual level 2 mV/m. Then you'd do:

IDL> idum=where(data.a.powradara gt 4.*(10.^(-6)),icount)

IDL> print,icount

1

There was only one event in June 1998 with this extremely high radar power threshhold specified; here is the raw spectrogram of the event. Note the incipient saturation due to the radar. (The saturation is analog, not ADC.) Note also that the event is actually triggered by the lightning, not the radar. (Three cheers for multichannel-coincidence triggering…)

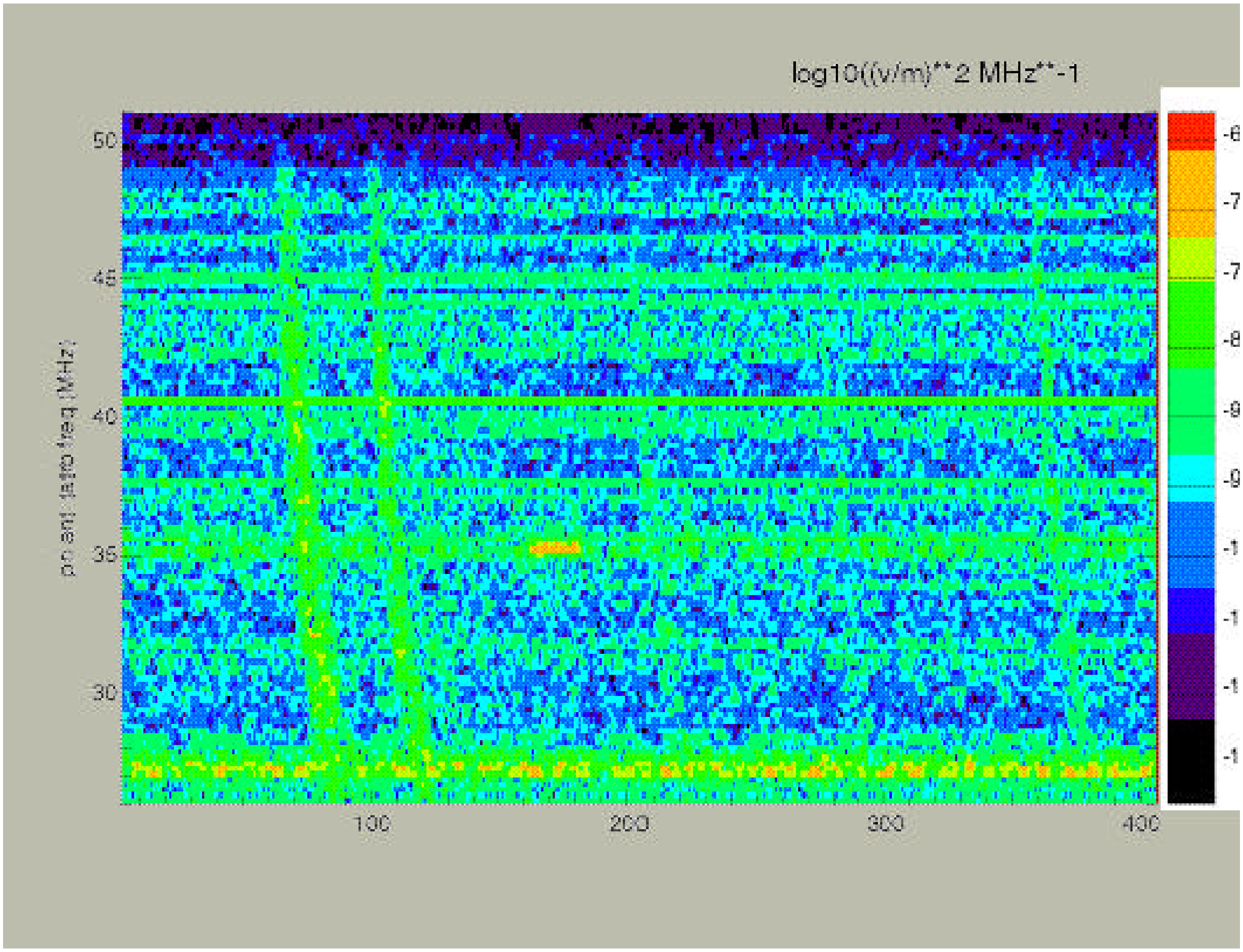


2.*(10.^(-7)),icount)

IDL> print,icount

26

which yielded 26 examples in June 1998. This is a collection of more modest radar pulses, one of which is the following:



(5b) Parse for coherent pulses

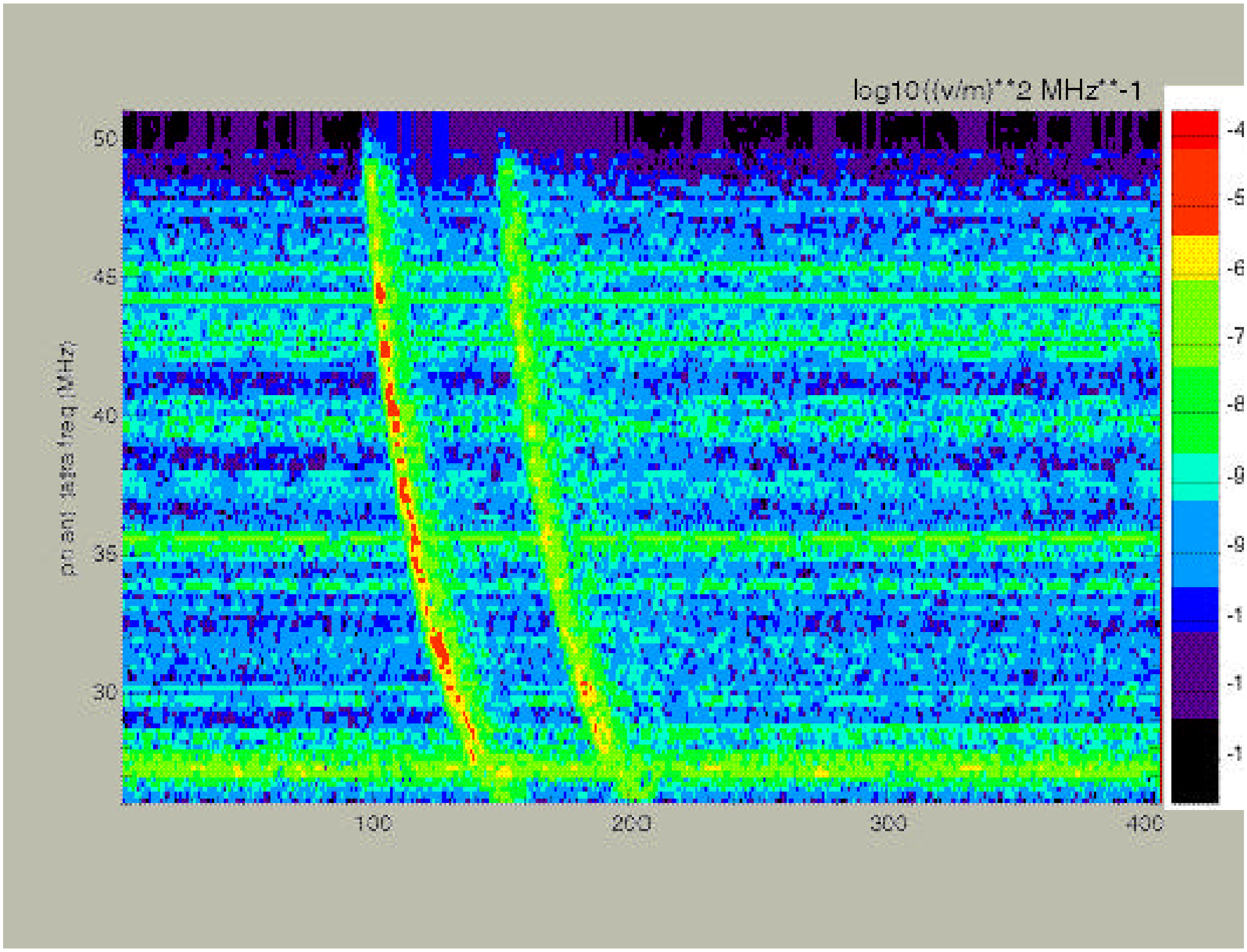
Suppose you seek very powerful lightning pulses which are TIPPs but coherent. There aren’t very many of these, but the few that exist can be found, e.g., as follows:

IDL> idum=where(data.a.tippsplita gt 0. and data.a.coherencea gt 0 and data.a.contrasta gt 200. and data.a.ppeaka gt (10.^(-6)),icount)

IDL> print,icount

7

There were 7 of these in June 1998. The raw spectrogram of one of them looks like:

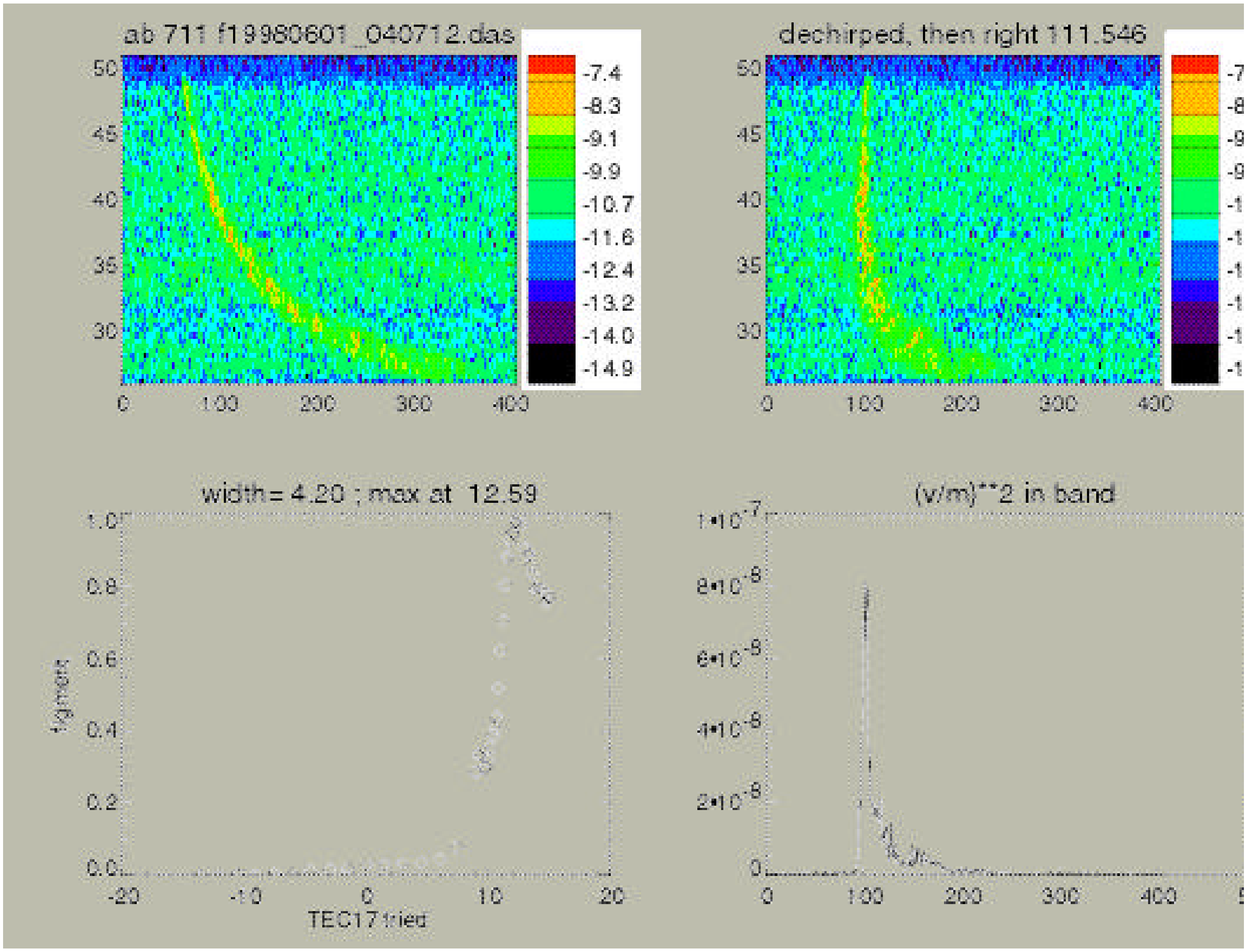


(5c) Parse for quartically distorted pulses in dual-freq data

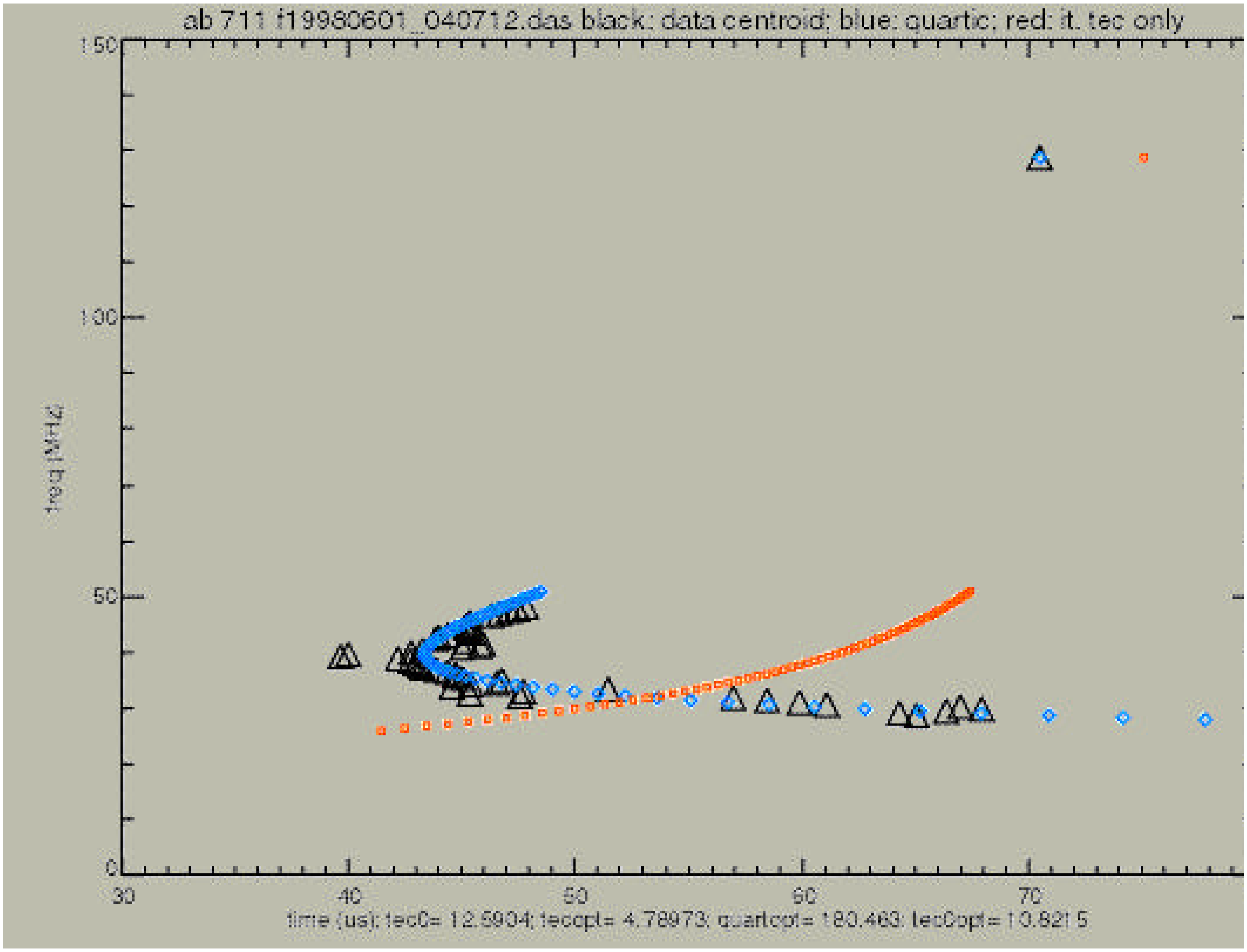
Suppose you want to find two-freq data which is extremely reliably fitted to a quartic, in which the quartic effect is very large, for which both bands contribute in the fit, and for which the pulse is high-contrast. Your selection might go as follows:

IDL> idum=where(data.l.bifreq gt 0 and data.l.sigratl gt 5. and data.l.quart100l gt 0. and data.l.tec0l gt 10. and data.l.contrastl gt 20.,icount)

The data for one such event is shown here (from other software):



The coordinated quadratic/quartic fit for this event is shown (in other software) as follows:



The two-freq data for power centroids is in black, the tec-only is in red, and the coordinated quadratic/quartic fit is in blue. The data has been pre-corrected for dispersion by a low-band-only TEC estimate. Inclusion of a coordinated quadratic/quartic fit has reduced the TEC estimate from 12.59 down to 4.79 ($X10^{17}$ m^{-2}).